Correlation-Based Interpolation of NSCAT Wind Data

Paulo S. Polito
Jet Propulsion Laboratory
California Institute of Technology

W. Timothy Liu
Jet Propulsion Laboratory
California Institute of Technology

Wenqing Tang
Jet Propulsion Laboratory
California Institute of Technology

Corresponding author: Paulo S. Polito
JPL/Caltech MS 300-323
4800 Oak Grove Dr.,
Pasadena, CA 91109, U.S.A.
Phone: 1-818-393-3839
Fax: 1-818-393-6720
Email: polito@pacific.jpl.nasa.gov

Abstract

Daily NASA Scatterometer (NSCAT) wind estimates cover about 75% of the globe. The remaining data gaps, which require interpolation, are regularly distributed in space and time. The development of this interpolation algorithm was guided by a balance between the smoothness of the end product and its fidelity to the original data. Three-dimensional matrices of autocorrelation coefficients incorporate information about the dominant propagation pattern into the interpolation program. These coefficients are continuously updated in space and time and are used as weights to interpolate each point in a regular space-time grid. For the first step, European Center for Medium-range Weather Forecast (ECMWF) wind data are used to simulate the NSCAT data distribution and interpolated using two different methods: one uses a single set of coefficients from a prescribed function based on the average decorrelation scales and the other uses the locally estimated autocorrelation coefficients. The comparison of these results with the original ECMWF maps favors those based on the autocorrelation. For the second step, daily maps of bin-averaged NSCAT wind data are compared to those interpolated by the correlation-based method and to those interpolated by successive corrections. Average differences between the original and interpolated fields are presented for the areas covered by the swath and for the gaps. The two-dimensional wavenumber spectra are also compared. The correlation-based interpolation (CBI) method retains relatively more small scale signal while significantly reducing the swath signature.

1. Introduction

The major advantage of using wind fields from satellite-based instruments is that global coverage is provided with a relatively fine resolution both in time and space. Although numerical weather prediction models provide global coverage, the spatial resolution of space-borne scatterometers is finer (Liu et al. 1998). Satellite scatterometers collect the backscattered radar signal from the ocean surface at an off-Nadir angle as they move along a polar orbit. The collected data are distributed in bands (swaths) that follow the orbit of the satellite. In one day NSCAT samples approximately 75% of the surface of the planet. Atmospheric features change shape and move while the satellite is still completing its daily sampling routine. Consequently a composite map of the wind samples has two major problems: regularly spaced data gaps and a regular pattern of ascending and descending orbits. The data gaps are regularly spaced in each map, but do not occur necessarily in the same place in successive maps. The orbit pattern is a consequence of averaging in a single map data from two or more satellite passes, collected at different times. The wind field changes from orbit to orbit and these changes become apparent due to strong wind gradients where the samples overlap.

The simplest solution would be to average over an area and/or a period long enough so that a complete, smooth global map is obtained. However, several problems arise from this approach. Small-scale atmospheric phenomena are generally short-lived and therefore lost in the averaging process. Averaging for a period that is long in relation to the motion of atmospheric features results in maps with weak winds and deformed fronts. More sophisticated techniques (Legler and O'Brien 1985; Kelly and Caruso 1990; Tang and Liu 1996) have demonstrated that these problems can be minimized by interpolating the gaps on global

maps assembled over a period on the order of one day. Optimal interpolation is generally performed using an analytical approximation of the autocorrelation fields, constant in space and time to minimize the computational cost of the procedure. The basic assumption in this case is that the statistics are approximately constant over the area and period covered by the interpolation. This assumption becomes a problem when a global wind data set is to be processed: for example, atmospheric features in the Tropical Pacific move and evolve in spatial and temporal scales different from those of the Antarctic Circumpolar region and wind patterns change significantly in time over the Indian Ocean during the Monsoon season.

The method presented in this study uses the three-dimensional autocorrelation coefficient matrices derived directly from daily bin-averaged maps in a regular grid for the interpolation onto the same grid. The autocorrelation matrices are continuously updated in space and time. The main objective of this study is to construct an interpolation scheme that **(a)** has NSCAT as the only source of data, therefore allowing for an independent comparison with other wind estimates, **(b)** minimizes the swath pattern signature significantly, and **(c)** has relatively small energy loss especially in the high-wavelength band.

The next section explains the interpolation method, its parameters and the general differences between this method and some other popular algorithms. Sections 3*a* and 3*b* discuss the application of this method in two different scenarios. In the first, experiments with simulated data show how changes in the input parameters affect the results. A general comparison is also made between the correlation-based interpolation and a comparable scheme based on a symmetric analytical correlation function. In the second, results from the interpolation of NSCAT data by two different schemes (correlation-based and successive corrections) are compared. As the two methods depend on different sets of parameters this comparison is not

3

intended to rank the methods. The goal is simply to build confidence in this new method by comparison to a well established one. A general discussion of the main results and the pertinent conclusions comprise the Section 4.

2. Interpolation Method

The correlation-based interpolation (CBI) method uses a sequence of regularly gridded maps of a wind speed component as input. These maps may have data gaps, indicated by a numerical flag, which will be interpolated over. The land masses are indicated by a different flag so that no interpolation is performed over land. The data grid is rectangular, global and $0.5° \times 0.5° \times 1$ day. If $z(x, y, t)$ is the wind component being interpolated, the algorithm starts by estimating the autocorrelation coefficients in the volume centered at (x, y, t) limited within lags of $(x_l = x \pm X)$, $(y_l = y \pm Y)$, and $(t_l = t \pm T)$. The maximum zonal, meridional and temporal lags are set as $X = 4°$, $Y = 3°$ and $T = 2$ days. These parameters are set empirically to close the gaps left by the scatterometer with the least amount of data thus minimizing the computational cost. Maximum lags (X, Y, T) that are very large compared to the average decorrelation scales increase the computational cost and yield marginal improvement in the results since the added data are uncorrelated.

As the wind field is not statistically homogeneous, the autocorrelation matrices are updated as often as possible. In the present case this is done at every grid point. Figure 1 shows three examples of autocorrelation matrices estimated from daily ECMWF zonal wind data with 1° resolution. The top row shows the the estimates at 165.5°W, 24.5°S on November 21, 1996. The middle row is for the same date, but at 179.5°E, 35°S. The bottom row is

4

for the same location as the middle row but 5 days later. A comparison of corresponding plots of the top and middle rows shows that the correlations may differ significantly from place to place in the same date (in this case within less than 2000 km). Similarly, the middle and bottom rows show different correlations estimated at the same place with a 5 days time difference.

Many popular interpolation schemes use elliptically symmetric analytical functions based on the average data covariance and error structures. In Section $3a$ an interpolation scheme based on an analytical symmetrical correlation function (ASC) is used to represent this class of methods in comparison to the CBI method. These analytical functions decay with distance from the origin with a decay scale adjusted to match the observed decorrelation scale in each dimension (Carter and Robinson 1987). However, Fig. 1 shows patterns that are not circularly symmetric and have variable decorrelation scales. Slanted patterns in the zonal $\times$ temporal (top right) or meridional $\times$ temporal (bottom center) correlation indicate phase propagation of atmospheric features. These patterns change in time and space. Therefore, the estimated correlation cannot be properly substituted by any single analytical function. This is the basic reason that justifies the computational cost of estimating and updating the local correlations. In addition, interpolation schemes whose kernel is a least-squares fitting solution (*e.g.* objective analysis) are sometimes unstable because of the inversion of rank-deficient matrices. Besides being mathematically simpler, this weighted average algorithm is unconditionally stable.

The autocorrelation coefficients, $c(x_l, y_l, t_l)$, used in the interpolation are calculated at

each location (x, y, t) by :

$$c = \frac{\sum\limits_{x_l, y_l, t_l} z \cdot z_l}{\sqrt{e_0 \sum\limits_{x_l, y_l, t_l} z^2 \cdot z_l^2}} \tag{1}$$

$$e_0 = \sqrt{x_l^2 + y_l^2 + t_l^2} \text{ with } e_0(0, 0, 0) = 1 , \tag{2}$$

where $z = z(x, y, t)$ and $z_l = z(x + x_l, y + y_l, t + t_l)$ are elements of the scalar matrix to be interpolated (*e.g.*, u, the zonal component of the wind vector). z and z_l are separated by spatial lags x_l and y_l, and temporal lag t_l. The autocorrelation is divided by $e_0(x_l, y_l, t_l)$ (Eq. 2) to avoid the large statistical fluctuations that occur at large lags due to the small number of factors in the sums.

The rms error of the correlation, $\sigma_c(x_l, y_l, t_l)$, is estimated at each location (x, y, t) by:

$$\sigma_c = \sqrt{\sigma_z^2 \cdot c^2 \left(\frac{\sum z_l}{\sum z \cdot z_l} - \frac{\sum z}{\sum z^2}\right)^2 + \sigma_{z_l}^2 \cdot c^2 \left(\frac{\sum z}{\sum z \cdot z_l} - \frac{\sum z_l}{\sum z_l^2}\right)^2}, \quad \sigma_{c(0,0,0)} = \sqrt{\frac{|\sigma_z^2 \cdot c|}{2\sqrt{N}}}, \tag{3}$$

where $\sigma_z(x, y, t)$ and $\sigma_{z_l}(x + x_l, y + y_l, t + t_l)$ are relative errors of the measurements associated with the precision of the instrument. These are assumed to be constant in space and time and have a value of 20% for all presented cases. The sums, like in Eq. 1, are performed over all possible lags.

Each interpolated point $z'(x, y, t)$ is obtained from the weighted average of N_s selected points (the selection criteria are discussed next) using the autocorrelation coefficients (c) as weights:

$$z'(x, y, t) = \frac{\sum c \cdot z}{\sum c} . \tag{4}$$

The rms error of the interpolated data, $\sigma_{z'}(x, y, z)$, is estimated from:

$$\sigma_{z'} = \sqrt{\sigma_c^2 \left(\frac{\sum z \sum c - N_s \sum c \cdot z}{(\sum c)^2}\right)^2 + \sigma_z^2} , \tag{5}$$

where N_s is the number of selected points. The sums in Eqs. 4 and 5 are performed over N_s points. The interpolated data are not recursively used in the interpolation process.

The criteria to select data points and weights are in the averaging process have two adjustable parameters: the maximum number of points N_p and the minimum correlation threshold c_{min}. The autocorrelation matrix with $N = (2X+1)(2Y+1)(2T+1)$ points is sorted and up to N_p coefficients larger than c_{min} are used.

If less than a minimum number of points (n_p) remains after the selection criteria are applied, a canonical correlation matrix (c_c) is used instead of the estimated one (thus $N \geq N_p \geq N_s \geq n_p$). For all experiments n_p is set to 5. The coefficients of the canonical correlation are calculated by:

$$c_c(x_l, y_l, t_l) = \frac{1}{1 + \sqrt{2t_l^2 + 0.5x_l^2 + 0.5y_l^2}} \, ,$$

(6)

with t_l in days and x_l, y_l in degrees. This is a symmetric inverse-distance function that has decay scales similar to the average of those inferred from ECMWF winds and is used in less than 1% of the cases. In the ASC scheme Eq. 6 is used to provide a symmetric set of weights for every point, as opposed to Eq. 1 used for the CBI method.

The zero-lag coefficient, $c(0,0,0)$, is a parameter of the interpolation scheme. Instead of being obtained from Eqs. 1 or 6 it can be simply set to a prescribed value. $c(0,0,0)$ is the weight attributed to the data at grid point being interpolated. Therefore, the larger this parameter is compared to 1, the closer the interpolated data will be to the input data.

3. Results

a. Experiments with simulated NSCAT data

1) Description of the tests and of the input data

A complete global map without gaps is used as a standard to compare the performance of the interpolation methods. Such maps can be obtained either from interpolated *in situ* data or from numerical weather prediction models. Interpolated *in situ* maps have very low temporal and spatial resolution due to the lack of data. Although the wind data generated by numerical models have less small-scale structure than the scatterometer data (Liu et al. 1998) they are still the best possible alternative for this study. Input fields that resemble scatterometer data with its regular gaps and gradients due to overlapping swaths are constructed from these global maps. Two sets of tests are performed with the simulated scatterometer data.

At first, zonal 10 m wind estimates from ECMWF, with a resolution of $1° \times 1°$, from November 15 to December 4 of 1996 are used to simulate NSCAT data. Two ECMWF maps, one taken at 12Z and one at 00Z are interpolated to $0.5° \times 0.5°$ grid using a bicubic interpolator. For each period of 12 hours before and after 12Z the NSCAT swath data are bin-averaged making two composite maps. These composites are used simply as a flag to delete data from two consecutive ECMWF maps in the same places where the NSCAT has sampling gaps. The two maps of ECMWF data with gaps are bin-averaged to generate one daily map. These daily maps are used as input to the interpolation programs and are referred simply as *input*. This method simulates not only the gaps but also the gradients

8

at the edges of overlapping swaths that characterize scatterometer data. These gradients are included in the simulations because they influence the correlation and can potentially change the results. A second ensemble of tests was performed using input data without the overlapping swath gradients. The data gaps remaining after bin-averaging 24 hours of NSCAT data were simply removed from the ECMWF file for 12Z. The *original* map (*i.e.* the "true" wind field) is the ECMWF synoptic map for 12Z. This allows for a comparison of interpolators based on a synoptic field.

The tests used two algorithms, the CBI which uses the estimated autocorrelation, and the ASC representing the interpolation schemes based on analytical functions. The combination of two input data sets and two algorithms yields the four groups of tests shown in Table 1. Within each group three parameters of the interpolation were changed: N_p, c_{min} and $c(0,0,0)$ (c_0 in the table headings).

2)   An index for spectral comparison of the CBI and ASC

An index (I) is devised to quantify the effects of changing the interpolation parameters in terms of a balance between smoothing and eliminating the swath gradients. This index compares the average value of the 2D spatial power spectrum distribution (PSD) in two wavenumber-domain regimes, one mostly affected by smoothing and the other mostly affected by the presence of swath gradients. Figure 2 illustrates the origin of this index. The insets of Plates **a** and **b** show two hypothetical interpolated fields, one very smooth (**a**) and the other with clear swath gradients (**b**) obtained from the same original ECMWF synoptical data shown in the background. An 11×11 square Gaussian smoother is used to generate the smooth field in **a**, which has no swath gradients. The field with overlapping swath gradients

(b) is obtained following the procedure described in Section 3a1 where no smoothing is performed. Land masses are filled with zeros for the purpose of spectral estimation. The PSD (P) shown in Plates c and d is estimated by

$$P(k,l) = 10 * \log_{10} \left(\frac{(\mathcal{F}_{256,256}(\mathbf{u}))^2}{256 \times 256} \right), \tag{7}$$

where $\mathcal{F}$ is the spatial 2D Fourier transform of $\mathbf{u}(x, y, t)$, the zonal wind estimates inside the white box measuring 256×256 points indicated in Fig. 2, Plates a and b. The absolute difference between the PSD of the smooth field (a, inset) and the PSD of the background field is shown in Plate e. The approximate location of the spectral regions that change the most as a result of smoothing are marked by the dashed white rectangle, defining wavenumber-domain regime e. Similarly in Plate f the regions that change the most as a result of the inclusion of swath gradients (b, inset) are marked by the white dashed rectangles, defining wavenumber-domain regime f. Tests performed by varying the amount of smoothing and the intensity of the gradients produce different spectra. Nevertheless, the white dashed wavenumber-domain regimes shown in Fig. 2 include representative samples of the spectral changes related to smoothing and swath gradients for all cases tested.

The index I is defined as:

$$\begin{cases} 10 \log_{10}(R_f) & = & < |P(\mathbf{u}_o) - P(\mathbf{u}_i)| >_f \\ 10 \log_{10}(N_f) & = & \dfrac{< P(\mathbf{u}_o) >_f}{< P(\mathbf{u}_o) >} \\ 10 \log_{10}(R_e) & = & < |P(\mathbf{u}_o) - P(\mathbf{u}_i)| >_e \\ 10 \log_{10}(N_e) & = & \dfrac{< P(\mathbf{u}_o) >_e}{< P(\mathbf{u}_o) >} \end{cases} \qquad I = \frac{\left| \dfrac{R_f}{N_f} - 1 \right|}{\left| \dfrac{R_e}{N_e} - 1 \right|} \tag{8}$$

where $\mathbf{u}_o$ is the original field without holes or gradients and $\mathbf{u}_i$ is an interpolated field. Angle brackets indicate an average over the wavenumber-domain regime indicated by the subscript.

The closer the PSDs of the interpolated and original fields are inside the wavenumber-domain regime f the smaller is the numerator of I. Similarly for wavenumber-domain regime e, the closer the PSDs the smaller the denominator of I. Therefore, if in Eq. 8 $I \gg 1$ the swath gradients are too intense for the amount of smoothing performed by this hypothetical interpolator. Conversely if $I \simeq 0$ this interpolated field is too smooth for the amount of gradients left.

3)   Comparison of the results from CBI and ASC

The skill of interpolation schemes is generally measured by their capability to reproduce the original field. This capability can be quantified in this particular set of tests by direct comparison with the original data (ECMWF synoptic field without holes or swath gradients). The input fields are obtained from the original data according to the procedure described in Section $3a1$. These fields provide data with or without gradients, only along the swath of the scatterometer. One can anticipate smaller discrepancies in the swaths in relation to areas with data gaps, therefore these areas (swath and gaps) are analyzed separately. Three statistical parameters were calculated: $\overline{\Delta u_s}$, the absolute relative difference, $|(u_o - u_i)/u_o|$, averaged over all the swaths covered by the input data; $\overline{\Delta u_h}$, averaged over data gaps, and $\overline{\Delta u_a}$, globally averaged. In addition to these parameters the globally average rms error $\overline{\sigma_a}$ (Eq. 5) relative to the local u_o was estimated. Table 1 presents the results for four groups of tests which used (i) the CBI method and input files with gradients, (ii) the CBI method and input files without gradients, (iii) the ASC method and input files with gradients and (iv) the ASC method and input files without gradients. In each group the parameters N_p, c_{min} and c_0 were changed. The index I is only relevant for tests performed using input files

11

with gradients.

A general comparison of the index I obtained for the CBI and ASC groups favors the CBI (*i.e.* I is closer to 1). The exceptions are extreme cases: tests 9 and 10 compared to tests 24 and 25. In these cases the number of degrees of freedom is reduced by setting a large weight (c_o) to the zero-lag factor. This effect can also be achieved, though not as effectively, by setting N_p to a small number or increasing c_{min} towards 1. As a consequence the value of I increases, indicating a relative excess of swath gradients in relation to the smoothness. The opposite effect is noticeable in tests 4 and 11 (or 19 and 26) where the number of averaged points is relatively large.

In tests that use a relatively small number of points both algorithms have the same performance in terms of $\overline{\Delta u_s}$. Otherwise the differences are small and favor the CBI. The relative discrepancy over data gaps ($\overline{\Delta u_h}$) is smaller for the CBI in all cases tested. Not surprisingly, $\overline{\Delta u_a}$ is also smaller for the CBI experiments.

The average relative rms error is strongly influenced by the number of points used in the interpolation. It is larger over data gaps compared to swaths because of the data availability. Table 1 shows only the global average ($\overline{\sigma_a}$) which is generally slightly larger (1% or less) for the CBI experiments.

The last four columns of Table 1 refer to experiments performed using input files without swath gradients and with data gaps. Similarly to the previous sets of experiments (with gradients) smaller differences favor the CBI method. The relatively small $\overline{\Delta u_s}$ in cases 9 and 10 (24 and 25) show the effect of increasing c_0. Contrariwise experiments 4 and 5 (19 and 20) show large discrepancies associated with increasing N_p. Changes in c_{min} have a lesser effect than changes in the other two parameters. Tests (not shown) performed with c_{min}

between zero and 0.2 give virtually identical results while using a c_{min} above 0.4 results in frequent gaps in the output field.

In Fig. 3 Plate **d** shows the input data for November 19, 1996. Plates **e** to **p** show the absolute differences between the input and interpolated fields for the CBI and ASC tests using input data with gaps and swath gradients. Relatively large errors tend to occur in areas with large data gaps such as the one that extends from the upper right corner to the SSE close to the Australian coast in Plate **d**. Furthermore the largest errors inside these data poor areas are due to changing winds: the blue spot between Australia and New Caledonia (Plates **e** to **p**) can be traced back to wind changes depicted in Plate **c**; the large error region to the SE of Japan near 32°N, 145°W appears in both Plates **a** and **c**.

Plates **e** to **p** present results from 6 pairs of tests comparing results from the CBI and ASC interpolation algorithms. As a general observation, the error structures for the CBI tests (**e**, **g**, **i**, **k**, **m**, and **o**) are smaller and slightly less intense than for the ASC tests (**f**, **h**, **j**, **l**, **n**, and **p**). Plates **e** and **f** represent a case in which N_p is relatively small. Results for **e** and **f** are almost identical and very small over swath areas (colored areas in Plate **d**). The clear contrast between the errors in swath areas compared to data gaps is an indication that the interpolated fields potentially have large swath gradients (confirmed by I in Table 1). Plates **g** and **h** were produced using the same parameters as **e** and **f** except that N_p is 20 times larger so the result is smoother. In this case the difference between the interpolation methods is more apparent than for **e** and **f**. The regions showing larger discrepancies are broader and more intense (green shades) in Plate **h** (ASC) compared to their counterparts in Plate **g** (CBI). The N_p and c_o of Plates **i** and **j** were 5 times larger than in **e** and **f**. These changes tend to reduce the errors compared to **g** and **h** while the difference between the interpolation

13

algorithms remain apparent. Plates **k** and **l** were obtained using the same parameters as **i** and **j** except for c_o which is larger by a factor of 10. The average error decreases compared to **i** and **j** for both the CBI and ASC tests while the difference is relatively clearer. In Plate **k** the larger errors (greens) are confined to the data gaps while in Plate **l** the large errors tend to invade part of the swath areas. Plates **m** and **n** were obtained using the same parameters as **i** and **j** except for c_o, which is 5 times smaller without changing the results significantly. Plates **o** and **p** were obtained using c_{min} set to 40%, with the other parameters set to the same values as **m** and **n**. In this case the number of points used in the interpolation is reduced. However, as only the weakly correlated points are removed, there is no noticeable increase in the error.

b. *Experiments with NSCAT data*

1) Description of the tests and of the input data

Daily bin-averaged global fields of the zonal component of the wind speed vector from the NSCAT data record are interpolated by the CBI method with the same parameter set used in experiment 6 (Table 1) and compared to a similar field obtained from successive corrections (SCI) (Tang and Liu 1996). The choice of parameter set is based on the value of I which is the closest to 1; likewise the errors are relatively low and none of its parameters were extreme.

The SCI scheme is based on a symmetric analytical basis function, conceptually similar to the ASC interpolation scheme. The most relevant characteristic of the SCI scheme that differentiates it from either the CBI or the ASC algorithms is that it is based on individual

14

passes as opposed to bin-averaged daily maps.

The experiments using simulated NSCAT data (Section 3a) were compared in a straight-forward way since the basis for comparison (*i.e.* a "true" wind field) is the original map from ECMWF, without holes or gradients. In the NSCAT case the crucial difference is the absence of a basis for comparison (*i.e.* there is no "true" wind field). The presented results compare one method relative to the other and can only provide a limited set of conclusions.

Figure 4 shows the spectra for both CBI and SCI estimated over a large area of the Pacific in November 19, 1996. The interpolated fields are shown as insets in Plates **a** and **b** (top row). Plate **b** (SCI) presents more small scale structures and more swath gradients than Plate **a** (CBI). However, it is necessary to quantify these differences and relate them to the bin-averaged NSCAT data. The spectra in Plates **c** and **d** show more high-wavenumber energy for the SCI case (**d**). Similarly to Fig. 2, the dashed boxes associated with smoothing (**e**) and with the presence of swath gradients (**f**) are plotted over the difference of the spectra. Visual comparison of these Plates convey little information about the performance of the algorithms, a quantitative comparison is necessary.

2) An index for spectral comparison of the CBI and SCI

An index J is defined to quantify the balance between removal of the swath signature in relation to the smoothing of the field, similarly to the index I in Section 3a2:

$$
\begin{cases}
10\log_{10}(R_f) = \; <|P(\mathsf{u}_{CBI}) - P(\mathsf{u}_{SCI})| >_f \\[2mm]
10\log_{10}(N_f) = \dfrac{< P(\mathsf{u}_{CBI}) + P(\mathsf{u}_{SCI}) >_f}{< P(\mathsf{u}_{CBI}) + P(\mathsf{u}_{SCI}) >} \\[2mm]
10\log_{10}(R_e) = \; <|P(\mathsf{u}_{CBI}) - P(\mathsf{u}_{SCI})| >_e \\[2mm]
10\log_{10}(N_e) = \dfrac{< P(\mathsf{u}_{CBI}) + P(\mathsf{u}_{SCI}) >_e}{< P(\mathsf{u}_{CBI}) + P(\mathsf{u}_{SCI}) >}
\end{cases}
\qquad
J = \dfrac{\left| \dfrac{R_f}{N_f} - 1 \right|}{\left| \dfrac{R_e}{N_e} - 1 \right|}
\qquad (9)
$$

15

The more similar the interpolated fields are, the more J approximates 1. If J is less than 1 then, compared to the SCI spectrum, the CBI has an excess of power in the region associated to smoothing (wavenumber-domain regime e in Plate e, Fig. 4) relative to the region associated with swath gradients (wavenumber-domain regime f) (*i.e.* the CBI removes relatively less swath gradients for the amount of smoothing it performs). Conversely, an index J larger than 1 indicates that, relative to the CBI, the SCI spectrum has more power in the wavenumber-domain regime e relative to wavenumber-domain regime f (*i.e.* the SCI removes relatively less swath gradients for the amount of smoothing it does).

3) Comparison of the results from CBI and SCI

For an 18-day period in which the interpolation was performed with the two algorithms, the distribution of the index is approximately lognormal with mean and standard deviation $J = 3.5 \pm 0.8$. This indicates that according to the criterium shown above the CBI algorithm provides more gradient removal relative to the amount of smoothing performed.

The average relative absolute differences between the interpolated fields and the input field are:

$$\overline{\Delta u_{CBI}} = \frac{2 < |u_{CBI} - u_o| >}{< |u_o| + |u_{CBI}| >}. \qquad \text{Similarly,} \qquad \overline{\Delta u_{SCI}} = \frac{2 < |u_{SCI} - u_o| >}{< |u_o| + |u_{SCI}| >}. \qquad (10)$$

The average relative absolute difference between the interpolated fields themselves is:

$$\overline{\Delta u_2} = \frac{2 < |u_{CBI} - u_{SCI}| >}{< |u_{CBI}| + |u_{SCI}| >}. \qquad (11)$$

The differences in Eq. 10 can only be calculated and averaged over the swath where the bin-averaged NSCAT data are available. The mean and standard deviations calculated over 18 realizations of the two algorithms are $\overline{\Delta u_{CBI_s}} = 0.13 \pm 0.01$ and $\overline{\Delta u_{SCI_s}} = 0.33 \pm 0.02$,

favoring the CBI algorithm. The parameters of the CBI algorithm (particularly c_0) can be adjusted to further reduce this difference (see Table 1), though no deliberate effort was made in this sense.

Equation 11 can be applied to swaths and interpolated data gaps alike. Using the same subscripts as in Section 3a3, the mean and standard deviation calculated using 18 realizations are $\overline{\Delta u_{2_a}} = 0.36 \pm 0.03$ (overall), $\overline{\Delta u_{2_s}} = 0.30 \pm 0.02$ (swaths), and $\overline{\Delta u_{2_h}} = 0.47 \pm 0.03$ (data gaps). Even though the Plates **a** and **b** in Fig. 4 have similar patterns, there are significant differences in the local value of the zonal wind, particularly in data gaps. Part of these differences is due to the design of the algorithms. The SCI method takes into account the time information in each satellite pass while the CBI method, based on the daily bin-averaged maps, treats all passes within one day as having the same time.

4. Conclusions

Four groups of experiments are performed with simulated NSCAT data derived from ECMWF zonal wind fields. These experiments apply the CBI and ASC algorithms to input files simulating NSCAT data with and without overlapping swath gradients. A spectral index I (Table 1) provides a quantitative measure of smoothing in relation to the reduction of swath gradients. Results indicate that in terms of I the CBI method achieves a better balance between smoothing and reduction of swath gradients than the ASC method. The discrepancies between the original and interpolated data are mostly due to the changes in local winds, as indicated by Fig. 3 and Table 1. These discrepancies are smaller for the CBI method than for the ASC.

17

The better performance of the CBI method results from the information about propagation and deformation of atmospheric features that enters the algorithm through the autocorrelation matrices. The ASC algorithm always assumes the same basic structure in space and time, as do most interpolation schemes.

These tests are performed based on numerical weather prediction winds which are characteristically smoother than scatterometer winds. The interpolation errors in these tests are probably underestimated in relation to those based on actual scatterometer winds.

Although limited by the differing assumptions of the CBI and SCI methods, a direct comparison between the interpolated and the bin-averaged NSCAT input data indicates that the CBI data is closer to the bin-averaged field than the SCI. Similarly to the previous experiments, an index J is defined to quantify smoothing in relation to the reduction of swath gradients. According to this index the CBI method performs better.

The fundamental difference between the CBI method and both the ASC and SCI is that the weight distribution (or basis function) of the CBI method is neither fixed nor symmetric. Instead, it is updated according to the local statistics. This characteristic allows the method to adjust itself to the continuous changes in the wind field and has a positive impact in the results.

5. Acknowledgements

References

Carter, E. F. and A. R. Robinson, 1987: Analysis methods for the estimation of oceanic fields. *Journal of Atmospheric and Oceanic Technology*, **1**, 49–74.

Kelly, K. and M. Caruso, 1990: A modified objective mapping technique for scatterometer wind data. *Journal of Geophysical Research*, **95**, 13483–13496.

Legler, D. M. and J. J. O'Brien, 1985: Development and testing of a simple assimilation technique to derive average wind fields from simulated scatterometer data. *Monthly Weather Review*, **10**, 1791–1800.

Liu, W. T., W. Tang, and P. S. Polito, 1998: NASA scatterometer provides global ocean-surface wind fields with more structure than numerical weather prediction. *Geophysical Research Letters*, **25**, 761–764.

Tang, W. and W. T. Liu, 1996: Objective interpolation of scatterometer winds. Technical Report 96-19, Jet Propulsion Laboratory, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA. 16 pp.

Table 1: Results using simulated NSCAT data: The columns show the experiment identification number with subscripts cross-referencing Fig. 3; the interpolation parameters N_p, c_{min} and c_0; the spectral index I (Eq. 8); and the interpolation discrepancies averaged over the swath, holes and global, $\overline{\Delta u_s}$, $\overline{\Delta u_h}$, and $\overline{\Delta u_a}$; $\overline{\sigma_a}$ is the overall mean statistical error. Experiments 1 to 15 refer to the CBI method; 16 to 30 refer to the ASC method. Columns 5 to 9 refer to the results obtained using input data with swath gradients; 10 to 13 used input data without gradients.

165.5°W, 24.5°S, 21/11/96

179.5°E, 35°S, 21/11/96

179.5°E, 35°S, 26/11/96